

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2002年12月27日

出願番号

Application Number:

特願2002-378956

[ST.10/C]:

[JP2002-378956]

出願人

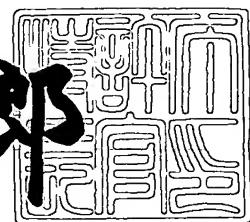
Applicant(s):

株式会社日立製作所

2003年 5月13日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3035591

【書類名】 特許願
【整理番号】 H02016111A
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 15/16
【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所 システム開発研究所内
【氏名】 細谷 瞳
【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社 日立製作所
【代理人】
【識別番号】 100075096
【弁理士】
【氏名又は名称】 作田 康夫
【電話番号】 03-3212-1111
【手数料の表示】
【予納台帳番号】 013088
【納付金額】 21,000円
【提出物件の目録】
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 高可用ディスク制御装置とその障害処理方法及び高可用ディスクサブシステム

【特許請求の範囲】

【請求項1】

ホストコンピュータとのインターフェースを有する複数のホストインターフェース部と、記憶装置とのインターフェースを有する複数のディスクインターフェース部と、前記記憶装置に対しリードまたはライトされるデータを一時的に格納する複数のキャッシュメモリ部を有し、前記複数のホストインターフェース部の各々は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記複数のディスクインターフェース部の各々は、前記記憶装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスク制御装置であって、

前記複数のホストインターフェース部と前記キャッシュメモリ部、及び、前記複数のディスクインターフェース部と前記キャッシュメモリ部との間が1つ以上のスイッチで構成されるスイッチ網を介して接続され、前記ホストインターフェース部、前記ディスクインターフェース部及び前記キャッシュメモリ部は、前記スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、前記スイッチは前記スイッチ網内の経路を前記IDを用いて指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を有し、さらに前記複数のキャッシュメモリ部は、該複数のキャッシュメモリ部における障害発生の有無を監視するための障害監視機構と前記スイッチ内のフォワーディングテーブルを制御するためのパス制御機構を有することを特徴とするディスク制御装置。

【請求項2】

前記パス制御機構は前記複数のキャッシュメモリ部における障害発生時に障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御することを特徴とする請求項1記載のディスク制御装置。

【請求項3】

ホストコンピュータとのインターフェースを有する複数のホストインターフェ

ース部と、記憶装置とのインターフェースを有する複数のディスクインターフェース部と、前記記憶装置に対しリードまたはライトされるデータを一時的に格納する複数のキャッシュメモリ部と、ホストインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する複数のリソース管理部を有し、前記複数のホストインターフェース部の各々は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記複数のディスクインターフェース部の各々は、前記記憶装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスク制御装置であって、

前記複数のホストインターフェース部、前記複数のディスクインターフェース部が1つ以上のスイッチで構成されるスイッチ網を介して前記キャッシュメモリ部と接続され、前記複数のホストインターフェース部、前記複数のディスクインターフェース部が前記スイッチ網を介して前記リソース管理部に接続され、前記ホストインターフェース部、前記ディスクインターフェース部、前記キャッシュメモリ部及び前記リソース管理部は、前記スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、前記スイッチは前記スイッチ網内の経路を前記IDを用いて指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を有し、さらに前記複数のキャッシュメモリ部と前記複数のリソース管理部は、その障害発生の有無を監視するための障害監視機構と前記スイッチ内の前記フォワーディングテーブルを制御するためのパス制御機構を有することを特徴とするディスク制御装置。

【請求項4】

前記パス制御機構は前記キャッシュメモリ部または前記リソース管理部における障害発生時に障害部位を回避するように前記スイッチ内の前記フォワーディングテーブルを制御することを特徴とする請求項3記載のディスク制御装置。

【請求項5】

前記キャッシュメモリ部における障害を監視するための前記障害監視機構が、前記キャッシュメモリ部の中にかえて、前記リソース管理部の中にあることを特徴とする請求項3または4記載のディスク制御装置。

【請求項6】

前記障害監視機構が、前記ホストインターフェース部及び前記ディスクインターフェース部でのリードまたはライトとともに動作することを特徴とする請求項1乃至5のいずれかに記載のディスク制御装置。

【請求項7】

前記キャッシュメモリ部における障害を処理するためのバス制御機構が、前記キャッシュメモリ部の中にかえて、前記リソース管理部に備わっていることを特徴とする請求項3乃至6のいずれかに記載のディスク制御装置。

【請求項8】

前記障害監視機構により障害を検出した際、前記バス制御機構により前記障害部位の前記ローカルIDと前記障害部位の機能を引き継ぐ前記交換部位の前記ローカルIDとを交換し、該ローカルID交換に対応して前記スイッチ網内経路を切り換えるよう前記スイッチ内フォワーディングテーブルを変更することを特徴とする請求項1乃至7のいずれかに記載のディスク制御装置。

【請求項9】

ホストコンピュータとのインターフェースを有する複数のホストインターフェース部と、記憶装置とのインターフェースを有する複数のディスクインターフェース部と、前記記憶装置に対しリードまたはライトされるデータを一時的に格納する複数のキャッシュメモリ部を有し、各ホストインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記記憶装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、

前記複数のホストインターフェース部と前記キャッシュメモリ部、及び、前記複数のディスクインターフェース部と前記キャッシュメモリ部との間が1つ以上のスイッチで構成されるスイッチ網を介して接続され、前記ホストインターフェース部、前記ディスクインターフェース部及び前記キャッシュメモリ部は、前記スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、前記スイッチは前記スイッチ網内の経路を前記IDを用いて指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を有し、さらに前記複数のキ

ヤッシュメモリ部は、障害監視機構とバス制御機構を有するディスク制御装置における前記キャッシュメモリ部の障害処理方法であって、

前記障害監視機構が前記キャッシュメモリ部における障害発生の有無を確認するステップと、該障害監視機構が該障害発生を前記バス制御機構に通知するステップと、該バス制御機構が前記通知された障害情報を解析するステップと、該バス制御機構が障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御するステップとを有することを特徴とする障害処理方法。

【請求項10】

ホストコンピュータとのインターフェースを有する複数のホストインターフェース部と、記憶装置とのインターフェースを有する複数のディスクインターフェース部と、前記記憶装置に対しリードまたはライトされるデータを一時的に格納する複数のキャッシュメモリ部と、ホストインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する複数のリソース管理部を有し、前記複数のホストインターフェース部の各々は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記複数のディスクインターフェース部の各々は、前記記憶装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、

前記複数のホストインターフェース部、前記複数のディスクインターフェース部が1つ以上のスイッチで構成されるスイッチ網を介して前記キャッシュメモリ部と接続され、前記複数のホストインターフェース部、前記複数のディスクインターフェース部が前記スイッチ網を介して前記リソース管理部に接続され、前記ホストインターフェース部、前記ディスクインターフェース部、前記キャッシュメモリ部及び前記リソース管理部は、前記スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、前記スイッチは前記スイッチ網での経路を前記IDを用いて指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を有し、さらに前記複数のキャッシュメモリ部と前記複数のリソース管理部は、障害監視機構とバス制御機構を有するディスク制御装置における前記キャッシュメモリ部および前記リソース管理部の障害処理方法であって、

前記障害監視機構が、前記キャッシュメモリ部または前記リソース管理部における障害発生の有無を確認するステップと、該障害監視機構が該障害発生を前記バス制御機構に通知するステップと、該バス制御機構が、前記通知された障害情報を解析するステップと、該バス制御機構が障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御するステップを有することを特徴とする障害処理方法。

【請求項11】

請求項10記載のディスク制御装置の前記キャッシュメモリ部における障害を監視するための前記障害監視機構が、前記キャッシュメモリ部の中にかえて、前記リソース管理部の中にあるディスク制御装置における障害処理方法であって、該障害監視機構が前記キャッシュメモリ部の障害発生の有無を確認するステップが、前記リソース管理部の中で実行されることを特徴とする障害処理方法。

【請求項12】

前記障害監視機構による障害発生の有無を確認するステップが、前記ホストインターフェース部及び前記ディスクインターフェース部でのリードまたはライト動作の際に実行されることを特徴とする請求項9乃至11のいずれかに記載の障害処理方法。

【請求項13】

請求項10乃至12のいずれかに記載のディスク制御装置の前記キャッシュメモリ部における障害を処理するためのバス制御機構が、前記キャッシュメモリ部の中にかえて、前記リソース管理部に備わっているディスク制御装置における障害処理方法であって、前記バス制御機構による前記フォワーディングテーブルを制御するステップが、前記リソース管理部の中で実行されることを特徴とする障害処理方法。

【請求項14】

前記バス制御機構により前記フォワーディングテーブルを制御するステップは、前記障害部位のローカルIDと障害部位の機能を引き継ぐ交換部位のローカルIDとを交換するステップと、該ローカルID交換に対応して前記スイッチ網内経路を切り換えるよう前記スイッチ内フォワーディングテーブルを変更するステップを

有することを特徴とする請求項9乃至13のいずれかに記載の障害処理方法。

【請求項15】

前記記憶装置は磁気ディスク装置であることを特徴とする請求項1乃至8のいずれかに記載のディスク制御装置。

【請求項16】

前記記憶装置は磁気ディスク装置であることを特徴とする請求項9乃至14のいずれかに記載の障害処理方法。

【請求項17】

複数のホストコンピュータと第1のネットワークを介して接続され、複数の磁気ディスク装置と第2のネットワークを介して接続されたディスク制御装置を備え、

上記ディスク制御装置は、

上記ホストコンピュータとのインターフェースを有する複数のホストインターフェース部と、

上記磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、

上記複数のホストインターフェース部及び上記複数のディスクインターフェース部との間が1つ以上のスイッチで構成されるスイッチ網を介して接続された複数のキャッシュメモリ部を備え、

前記複数のホストインターフェース部、前記複数のディスクインターフェース部及び前記複数のキャッシュメモリ部は、前記スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、

前記スイッチは、前記スイッチ網内の経路を前記IDを用いて指定するためのフォワーディングテーブルと、該テーブルを変更する変更手段を有し、

前記複数のキャッシュメモリ部は、

該複数のキャッシュメモリ部における障害発生の有無を監視するための障害監視機構と、

前記スイッチ内のフォワーディングテーブルを制御するためのバス制御機構を有することを特徴とするディスクアレイサブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、データを複数の磁気ディスク装置に格納するディスクシステム装置の制御装置に関する。

【0002】

【従来の技術】

企業間の電子商取引や社会基盤としての金融システムなどでは、高度な信頼性が要求され、その根幹をなす基幹ストレージ・システムに対しても、極めて高い可用性が求められている。これら基幹ストレージ・システムでは、その可用性を高めるために、内部を冗長構成とし、障害発生時には自動的に故障個所を切り離して、正常な冗長部分で動作を継続するための自動障害回復機能を備えたディスク制御装置が広く使われている。

【0003】

例えば、図9に示した、従来より知られているディスク制御装置は、ホストコンピュータ60との間のデータ転送を実行する複数のホストインターフェース部1Xと、磁気ディスク装置70との間のデータ転送を実行する複数のディスクインターフェース部2Xと、磁気ディスク装置70のデータを一時的に格納するキャッシュメモリ部3Xと、ディスク制御装置104に関する制御情報（例えば、ホストインターフェース部1X及びディスクインターフェース部2Xとキャッシュメモリ部3Xとの間のデータ転送制御に関する情報、磁気ディスク装置70に格納するデータの管理情報）を格納するリソース管理部5Xとを備えている。

ホストインターフェース部1X及びディスクインターフェース部2Xとキャッシュメモリ部3Xとの間は、データインターフェース信号6により接続される。ホストインターフェース部1Xとキャッシュメモリ部3Xとの間、及び、ディスクインターフェース部2Xとキャッシュメモリ部3Xとの間の接続にスイッチ4Xを用いることもある。ホストインターフェース部1X及びディスクインターフェース部2Xとリソース管理部5Xとの間は、管理インターフェース信号7により接続される。リソース管理部5Xと、ホストインターフェース部1X、及び、ディスク

インターフェース部2Xとの接続は、スイッチを介しても、介さなくても良い。これにより、リソース管理部5Xおよびキャッシュメモリ部3Xは全てのホストインターフェース部1X及びディスクインターフェース部2Xからアクセス可能な構成となる。

【0004】

図12に示すように、ホストインターフェース部1Xは、ホストインターフェース信号1との入出力を処理するチャネルプロトコル処理部90、データインターフェース信号6との入出力を処理するための内部プロトコル処理部8X、管理インターフェース信号7との入出力を処理するためのプロセッサインターフェース17、及びホストコンピュータ60に対する入出力を制御するプロセッサ14とローカルメモリ15を有している。

ディスクインターフェース部2Xも、構造としては、ホストインターフェース部と同様であるが、ホストインターフェース信号1の代わりにディスクインターフェース信号2がチャネルプロトコル処理部90に接続され、プロセッサ14においては、ホストインターフェース部で行われる制御に加えて、RAID機能の実行も行う。

ホストインターフェース部1X、及び、ディスクインターフェース部2Xが、キャッシュメモリ部3Xと通信を行う場合、データの先頭に宛先アドレスを付加したパケットを使用してパケット転送を行う。

ホストインターフェース部1X、もしくは、ディスクインターフェース部2X内のプロセッサ14の制御で生成されたパケットは、データインターフェース信号6を介して、スイッチ4Xに送られる。スイッチ4Xは、図10で示したように、データインターフェース信号6に接続した複数のバスインターフェース41Xと、パケットバッファ43とアドレスラッチ44とセレクタ48を備えている。バスインターフェース41X内には、パケットからアドレス情報を取り出すヘッダ解析部42Xが含まれており、それによって解析抽出されたパケットアドレスがアドレスラッチ44に取り込まれる。一方、送られてきたパケットは、バスインターフェース41Xを通して、パケットバッファ43に格納される。アドレスラッチ44からは、パケット宛先に応じたセレクタ制御信号47が生成され、パケ

ットバッファ43に格納されたパケットの送出先をセレクタ48によって切り換える。

【0005】

スイッチ4Xで、宛先ごとに振り分けられたパケットは、再び、データインターフェース信号6を介して、目的のキャッシングメモリ部3Xに転送される。キャッシングメモリ部3Xは、図11で示したように、データインターフェース信号6に接続した複数のデータバスインターフェース31Xと、パケットバッファ33と調停回路39とセレクタ38を備えている。データバスインターフェース31X内には、パケットからアドレス情報を取り出すヘッダ解析部32Xが含まれており、それによって解析抽出されたパケットアドレスは、調停回路39に取り込まれる。一方、送られてきたパケットは、バスインターフェース31Xを通して、パケットバッファ33に格納される。調停回路39は、複数のデータバスインターフェース31Xの中から、いずれかを選択し、その選択結果に応じたセレクタ制御信号を生成する。このセレクタ制御信号で、セレクタ38を切り換えることにより、所望のパケットバッファ33の内容をキャッシングメモリ37に、メモリ制御回路35を介して、書き込むことができる。パケットバッファ33に格納されたパケットがメモリ読み出しの要求であった場合、指定された領域のキャッシングメモリ37の内容を、上記と逆の経路をたどることによって、ホストインターフェース部1X、もしくは、ディスクインターフェース部2Xに返送する。

【0006】

ホストインターフェース部1X、及び、ディスクインターフェース部2Xが、リソース管理部5Xと通信を行う場合も、データインターフェース信号6の代わりに、管理インターフェース信号7が用いられる以外は、キャッシングメモリ部との通信と同様なパケット転送が行われる。リソース管理部5Xは、図11に示したキャッシングメモリ部と、インターフェース信号を除いて同等な構成になっている。

ここで、キャッシングメモリ部3X、及び、リソース管理部5Xは、複数のホストインターフェース部1X、ディスクインターフェース部2Xからアクセスされるシステム共通のリソースであり、その可用性がシステムの信頼性に大きな影響を与える。

るため、同等機能を複数備えた冗長構成となっており、たとえ、一方に障害が発生しても、残りの正常な部分で動作を継続できるように設計されている。具体的には、ホストインターフェース部1X、もしくは、ディスクインターフェース部2X内のいずれかのプロセッサ14が、複数あるキャッシュメモリ部3X、もしくは、リソース管理部5Xのいずれかの障害を検知した場合、障害を検知したプロセッサ制御で、障害部分を閉塞して残りのキャッシュメモリ部3X、もしくは、リソース管理部5Xに、その機能を引き継がせるとともに、他の全てのプロセッサ14に対して障害通知を行う。障害通知を受けた全てのプロセッサは、障害に伴うシステム構成・通信経路の変更処理をそれぞれ行うことによって、すべてのホストインターフェース部1Xとディスクインターフェース部2Xで、障害部分の切り離しを実現することが可能となる。

このように従来のディスク制御装置104では、キャッシュメモリ部3X、もしくは、リソース管理部5Xのような共通リソース障害に伴うシステム構成・通信経路の変更処理が、複数あるホストインターフェース1X、および、ディスクインターフェース2X内のプロセッサそれぞれで分散して行われていた。そのため、共通リソースの障害においては、分散配置されたプロセッサ間でのブロードキャスト通信を含む複雑な処理が必要であった。

【0007】

ディスク制御装置を高信頼化する別の従来例として、共有システムリソースとシステムリソース・クライアントとの間で可用性の高いネットワーク通信を提供する障害処理機構が提案されている（例えば、特許文献1参照。）。この従来例においても、さきに説明した従来例と同様、複数あるプロセッサごとに経路変更（ルーティングテーブル変更）を行う方式となっている。

【0008】

また、ディスク制御装置の高可用性を実現する別の従来例として、ホストコンピュータとディスクアレイサブセットとの間に配置して、両者の間のアドレス変換を行うスイッチを備えた記憶装置システムが提案されている（例えば、特許文献2参照。）。この従来例では、複数あるディスクアレイサブセットに障害が発生した場合、スイッチ内でパケットの解釈を行って障害部分への要求を同等の機

能を有した冗長部分の宛先に変更することで、経路変更などの障害処理を行う方
式となっている。

【特許文献1】

特開2002-41348号公報

【特許文献2】

特開2000-242434号公報

【0009】

【発明が解決しようとする課題】

キヤッシュュメモリ部、リソース管理部のような共通リソースの障害は、ストレージ・システム、ひいては、ホストコンピュータで実行されているアプリケーションの動作不良を引き起こすため、すみやかな回復処理が行われなければならない。しかし、図9、図10、図11、図12に示した従来技術では、すべてのホストインターフェース部1X、ディスクインターフェース部2Xにおいて経路変更が必要なため、障害処理に時間がかかり、ホストコンピュータとの間でのリード／ライトタスクの継続が行えずに、ストレージシステムの性能劣化やアプリケーション・プログラムの動作不良を引き起こす場合があった。また、この障害処理には、ホストインターフェース部1X、ディスクインターフェース部2Xに高機能プロセッサと複雑な制御プログラムが必要となり、製造コストの増大、信頼性の低下を招いていた。特許文献1に示した別の従来技術においても、複数あるプロセッサごとにルーティングテーブル変更を行う必要があり、同様の課題があった。

【0010】

また、特許文献2で開示されている別の従来技術では、パケット宛先変更機能を有するスイッチを導入することによって、複数あるディスクアレイ・サブセット間で機能を引き継がせる等の障害回避動作をスイッチ内処理のみで行うことが可能となる。しかし、その一方で、パケットごとに宛先の解釈が必要となり、障害発生時のみならず、通常動作時においても、処理に多くの時間を要し、ストレージシステムの性能劣化を招くという課題を有している。

【0011】

本発明の目的は、上記従来技術の欠点を改善し、障害発生時に、迅速かつ高信頼に障害処理を行い、かつ、通常動作時においても性能劣化を起こさない高可用のディスク制御装置、および、その障害処理方法を提供することにある。

より具体的には、本発明の目的は、システム共通リソースの障害時を含め、いかなる場合においても、ストレージ・システムの性能劣化や、ホスト・アプリケーションの動作不良を引き起こすことのない高可用性ディスク制御装置を提供することにある。

【0012】

【課題を解決するための手段】

上記目的を達成するため、本発明では、ホストコンピュータとのインターフェースを有する複数のホストインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、ホストインターフェース部及びディスクインターフェース部とキャッシュメモリ部との間のデータ転送に関する制御情報を格納するリソース管理部を有し、各ホストインターフェース部は、ホストコンピュータとのインターフェースとキャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、磁気ディスク装置とのインターフェースとキャッシュメモリ部との間のデータ転送を実行するディスク制御装置であって、

複数のホストインターフェース部とキャッシュメモリ部及び複数のディスクインターフェース部とキャッシュメモリ部との間が1つ以上のスイッチで構成されるスイッチ網を介して接続され、スイッチがスイッチ網内での経路を指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を有し、また、ホストインターフェース部、ディスクインターフェース部及びキャッシュメモリ部が、スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を有し、さらに複数のキャッシュメモリ部が、その障害発生の有無を相互に監視するための障害監視機構と障害発生時に障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御するためのパス制御機構を有するディスク制御装置を提供する。

【0013】

【発明の実施の形態】

以下、大容量のデータの記憶装置として磁気ディスク装置を例にとって説明するが、大容量記憶装置として磁気ディスクに限られるものではなく、例えばDVDのような大容量記憶装置であって良い。

本発明の実施の形態の1つとして、好ましくは、前記複数のホストインターフェース部と前記キャッシュメモリ部との間、及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間を、1つ以上のスイッチで構成されるスイッチ網を介して接続し、スイッチにスイッチ網内での経路を指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を設け、また、ホストインターフェース部、ディスクインターフェース部及びキャッシュメモリ部に、スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を設け、さらに複数のキャッシュメモリ部に、その障害発生の有無を監視するための障害監視機構と障害発生時に障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御するためのパス制御機構を設ける。

また、好ましくは、前記複数のホストインターフェース部、前記複数のディスクインターフェース部、前記キャッシュメモリ部、前記リソース管理部との間を1つ以上のスイッチで構成されるスイッチ網を介して接続し、スイッチにスイッチ網内での経路を指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を設け、また、ホストインターフェース部、ディスクインターフェース部及びキャッシュメモリ部に、スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を設け、さらに複数のキャッシュメモリ部に、その障害発生の有無を監視するための障害監視機構と障害発生時に障害部位を回避するよう前記スイッチ内のフォワーディングテーブルを制御するためのパス制御機構を設ける。

また、好ましくは、前記リソース管理部に、前記キャッシュメモリ部、もしくは、前記リソース管理部の障害を監視するための障害監視機構を設ける。

【0014】

また、好ましくは、前記ホストインターフェース部及び前記ディスクインター

フェース部に前記キャッシュメモリ部、もしくは、前記リソース管理部の障害監視機構に対して障害を報告する機能を設ける。

また、好ましくは、前記リソース管理部に、前記キャッシュメモリ部の障害を処理するためのバス制御機構を設ける。

その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【0015】

以下、本発明の実施例を図面を用いて説明する。

《実施例1》

図1、図2、図6、図7、及び図8に、本発明の一実施例を示す。

図2に示したディスク制御装置100は、ホストコンピュータ60とのインターフェース部（ホストインターフェース部）10と、磁気ディスク装置70とのインターフェース部（ディスクインターフェース部）20と、キャッシュメモリ部30、スイッチ40、リソース管理部50を有し、ホストインターフェース部10及びディスクインターフェース部20と、キャッシュメモリ部30及びリソース管理部50との間は、スイッチ40を介して、内部インターフェース信号4で接続されている。すなわち、内部インターフェース信号4を介して、全てのホストインターフェース部10、全てのディスクインターフェース部20から、全てのキャッシュメモリ部30、あるいはリソース管理部50へアクセス可能な構成となっている。

【0016】

図8に示すように、ホストインターフェース部10は、ホストインターフェース信号1との入出力を処理するチャネルプロトコル処理部90、データインターフェース信号との入出力を処理するための内部プロトコル処理部80を有し、ホストコンピュータ60とキャッシュメモリ部30間のデータの転送、及びリソース管理部50との間の制御情報の転送を実行する。

ディスクインターフェース部20も、構造としては、ホストインターフェース部と同様であるが、ホストインターフェース信号1の代わりにディスクインターフェース信号2を有し、磁気ディスク装置70とキャッシュメモリ部30間のデ

ータの転送、及びリソース管理部50との間の制御情報の転送を実行する。

キャッシュメモリ部30は、図7に示すように、内部インターフェース信号4との入出力処理を行う内部プロトコル処理部80とプロセッサ36とキャッシュメモリ37とメモリ制御回路35とDMAエンジン34を有し、磁気ディスク装置70へ記録するデータや磁気ディスク装置から読み出したデータを一時的に格納する。

【0017】

リソース管理部50も、キャッシュメモリ部30と同等の構成を有し、システム構成などの管理制御情報を維持する。

【0018】

スイッチ40は、図6に示すように、内部インターフェース信号4に接続した複数のバスインターフェース41と、パケットバッファ43とアドレスラッチ44とセレクタ48を備え、ホストインターフェース部10およびディスクインターフェース部20と、キャッシュメモリ部30およびリソース管理部40との間の経路接続を行う。

なお、可用性向上のために、ホストインターフェース部10、ディスクインターフェース部20、キャッシュメモリ部30、リソース管理部50を、それぞれ複数のポートを有する構成とし、スイッチ40との間にそれぞれ複数の転送経路を設けることも出来る。

【0019】

ホストインターフェース部10、ディスクインターフェース部20、キャッシュメモリ部30、リソース管理部50の内部プロトコル処理部80には、内部インターフェース信号4の接続しているスイッチ網内の宛先を一意に特定するためのローカルID(LID)を保持するLID情報81が備わっている。

【0020】

一方、スイッチ40内には、ポート番号(バスインターフェース41の位置)とLIDとの対応関係を示すフォワーディング・テーブル46が備わっている。フォワーディング・テーブル46の例を、図1(1)に示す。この例では、2つのホストインターフェース10と2つのディスクインターフェース20が2つの

スイッチ40A、40Bを介して、2つのキャッシュメモリ（共通リソース）30A、30Bに接続している。ホストインターフェース10、ディスクインターフェース20、キャッシュメモリ30A、30Bはそれぞれ2つずつの内部インターフェース信号とそれに対応したローカルID（LID）情報を持っている。スイッチ40A、40Bは、それぞれ8つのポート（バスインターフェース41）とそれに対応したポート番号を持っている。フォワーディング・テーブル46とは、このLIDとポート番号との対応表であって、例えば、スイッチ46Aのフォワーディング・テーブルAでは、LID①、③、⑤、⑦、⑨、(11)が、それぞれポートa、b、c、d、e、fと接続していることを示している。このフォワーディング・テーブルを参照することにより、パケット宛先（LID）によって、どのポートに送出すればよいかがわかる。

【0021】

内部インターフェース信号の接続しているスイッチ網は、例えば、キャッシュメモリ部30内のプロセッサ36で実行されるネットワーク管理プログラムによって、維持管理されている。ネットワーク内のLID情報81やスイッチ内のフォワーディング・テーブル46は、内部インターフェース信号4を介して、ネットワーク管理プログラムにより、設定更新される。

本発明のディスク制御装置における通常動作の一例として、図2、図6、図7、図8を用い、ホストコンピュータ60からディスク制御装置100を介して磁気ディスク装置70へ読み出し要求を発行する場合の動作について説明する。

まず、ホストコンピュータ60は、自身が接続されているホストインターフェース部10にデータの読み出し要求を発行する。要求を受けたホストインターフェース部10は、リソース管理部50にアクセスし、要求されたデータがどの磁気ディスク装置70内に格納されており、当該磁気ディスクがどのキャッシュメモリ部30に制御されているかを調べる。リソース管理部50には、要求データのアドレスからこれらの情報を検索するためのテーブルが格納されており、要求されたデータをもとに管轄のキャッシュメモリ部を調べることができる。次に、要求を受けたホストインターフェース部10では、要求データを管理しているキャッシュメモリ部30に、読み出し要求を転送する。キャッシュメモリ部30は

、キャッシュメモリ37に要求されたデータが格納されているかどうかを確認する。キャッシュメモリ部30内にデータが存在しなかった場合、プロセッサ36は、要求データを磁気ディスク装置70から読み出し、キャッシュメモリ37に格納する。キャッシュメモリ部30は、キャッシュメモリ37内に格納された要求データを、ホストインターフェース部10まで転送し、ホストコンピュータ60に送る。

ホストインターフェース部10及びディスクインターフェース部20がキャッシュメモリ部30、及びリソース管理部50とスイッチ40を介して通信する場合、パケット宛先にLID情報を用い、スイッチでの経路変更にフォワーディング・テーブル46を使用する。

例えば、ホストコンピュータ60からの読み出し要求を、スイッチ40を介して、キャッシュメモリ部30に転送する場合、ホストインターフェース部10では、ホストインターフェース信号1を、チャネルプロトコル処理部90と内部プロトコル処理部80を介して、内部インターフェース信号4に変換する。その際、読み出し要求パケットの宛先には、送り先のキャッシュメモリ部のLIDを設定する。ホストインターフェース部からの要求パケットを受けたスイッチ40では、経路変更手段45内のフォワーディング・テーブル46に従い、ヘッダ解析部42で抽出されたパケット宛先に応じたセレクタ制御信号47を生成し、パケットバッファ43に格納されたパケットの送出先をセレクタ48によって切り換えて、所望のキャッシュメモリ部30に向けてパケットを転送する（図16）。

【0022】

ディスクインターフェース部20が、キャッシュメモリ部と通信を行う場合、もしくは、ホストインターフェース部10、及び、ディスクインターフェース部20が、リソース管理部50と通信を行う場合も、ホストインターフェース部とキャッシュメモリ部との間の通信と同様なパケット転送が行われる。

なお、リソース管理部50と、ホストインターフェース部10およびディスクインターフェース部20との接続に使われているスイッチ40を含んだスイッチ網は、キャッシュメモリ部30とホストインターフェース部10およびディスクインターフェース部20との接続に使われているスイッチ網と同じであっても、別

の専用ネットワークであっても良い（図17）。また、スイッチ40を使わずに、直接接続する形態であっても構わない。

次に、本発明のディスク制御装置の特徴である障害回復動作の一例として、図1、図2を用い、2つのキャッシュメモリ部30間での障害監視機構とバス制御機構の動作について説明する。

キャッシュメモリ部30は、その可用性向上のため、同等機能を有するマスタとスレーブの2つのキャッシュメモリ部を有している。スレーブキャッシュメモリは、マスタキャッシュメモリに障害が発生した際に、その機能を引き継ぐためのもので、ホットスタンバイで動作している。これらマスタ・キャッシュメモリ部とスレーブ・キャッシュメモリ部は、スイッチ40を介して、相互にその動作を確認するための障害監視機構Cを有している。すなわち、一定間隔ごとに自分の動作状況を報告するパケットを生成し、お互い相手の動作状態を監視することができる。その動作の概要を図13に示す。障害監視機構では通信が行われる度に、正常なシーケンスとACKがチェックされ、異常が発見された場合には、主経路もしくは副経路を介して、バス制御機構に障害通知を行う。図2に示した構成では、マスタおよびスレーブキャッシュメモリ部が、それぞれバス制御機構を有しているので、そのいずれかのキャッシュメモリ部に異常が発生した場合、その障害情報は、ただちにもう一方のキャッシュメモリ部で検知することができる。障害を検知したキャッシュメモリ部では、障害を起こしているキャッシュメモリ部を閉塞するとともに、内部に備えたバス制御機構Pによって、システム構成を変更し、ホストインターフェース部10およびディスクインターフェース部20が、障害キャッシュメモリ部にアクセスしないように制御する。

以下、バス制御機構Pの仕組みを、図14を使って説明する。障害監視機構から障害通知を受けたバス制御機構は、その通知の妥当性を確認した後、可用性向上のため複数個設けてあるバス制御機構間で、障害情報の同期化を行う。その後、障害解析を行い、現時点で障害特定が可能かどうか判断する。特定が出来ない場合、特定可能になるまで障害処理を遅延させる。特定可能になった場合、障害知情報によりアクセス障害なのか機能障害なのかを判断し、機能障害の場合、障

害部位を含んだ冗長部位間で、処理途中のJ O Bなどの同期化処理を試みる。その後、障害部位に対するアクセス・パスを冗長部位に振り替える交替バス処理を行う。図1で具体的に説明すると、マスタキャッシュメモリ部で機能障害が発生しスレーブキャッシュメモリ部にフェールオーバーする場合は、以下のようになる。障害のない正常動作状態における、ホストインターフェース部10、ディスクインターフェース部20、キャッシュメモリ部30A、30B、スイッチ40A、40BのL I Dとフォワーディング・テーブルの値は、図1(1)に示した通りである。ここで、マスタ・キャッシュメモリ部30Aに障害が発生し、スレーブ・キャッシュメモリ部30Bの障害監視機構によって、当該障害が検出された場合、30Bのバス制御機構Pにより、30Aの機能を30Bが引き継ぐとともに、30A宛のパケットを30Bに振り替える制御を行う。すなわち、30Aの2つのL I D(⑨、(10))と30Bの2つのL I D((11)、(12))とを交換し、それに対応して、フォワーディング・テーブル46Aと46Bを変更する。その結果、L I Dとフォワーディングテーブルは図1(2)のようになり、30Aへのアクセスは、すべて30Bに振り向けられて、障害部位30Aのシステムからの切り離しが完了する。30Aの動作を30Bが引き継ぐには、30Aと30Bの内容が一致している必要があるが、それらは、通常の同期動作により実現される。すなわち、常に30Aと30Bが同じ内容になるよう、両方に対して同じアクセス要求を発生させる方法や、定期的に両者でデータをコピーする方法が考えられる。

なお、リソース管理部50も同様の障害監視機構Cとバス制御機構Pを備え、同様の手順で障害回復動作を行うことができる。これら障害監視機構やバス制御機構は、キャッシュメモリ部30、もしくは、リソース管理部50内のプロセッサ36によって実行される制御プログラムとして実現することができる。また、本実施例では、リソース管理部50を独立に設けた構成としているが、リソース管理部をキャッシュメモリ部30の中に設けることもできる。また、図1で、スレーブキャッシュメモリ部内のL I Dとマスタキャッシュメモリ部内のL I Dを交換するのではなく、スレーブ側でマスタ側のL I Dを追加して持つようにすることも可能である。その場合、障害前のスレーブ側L I Dが障害後も有効になると

いうメリットがある。

【0023】

本実施例によれば、キャッシュメモリ部30もしくはリソース管理部50の障害に対して、スイッチ40のフォワーディング・テーブルとキャッシュメモリ部30もしくはリソース管理部50のLIDの変更のみで、障害部位の切り離しが完了し、従来技術のように、複数のホストインターフェース部10及びディスクインターフェース部20の間でのブロードキャスト通信や複雑な制御は不要である。そのため、障害発生時においても、迅速、かつ、高信頼な障害回復処理を実現することができ、ストレージシステムの性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことがない。

また、本実施例のスイッチ内フォワーディング・テーブルは、障害発生時にのみ変更されものであって、従来技術のように、通信を行うたびにパケットの宛先の解釈や変更が行われる複雑なスイッチを必要としない。そのため、障害のない通常の動作においての性能劣化は皆無で、かつ、低コストで高信頼に製造することが可能である。

【0024】

《実施例2》

図3に、本発明の他の実施例を示す。

図3に示した実施例は、キャッシュメモリ部30、及び、リソース管理部50が、それぞれ障害通知のための専用線であるハートビート信号3を備えていることと、内部インターフェース信号のスイッチ網が多段のスイッチ40で構成されていることを除いて、実施例1の図2に示す構成と同様である。ただし、キャッシュメモリ部30は、障害監視機構Cのみを備え、バス制御機構は実装されていない。リソース管理部50は、障害監視機構C、バス制御機構Pともに備えている。

キャッシュメモリ部、および、リソース管理部は、マスタとスレーブにより冗長構成を形成しており、基本的には同じデータ内容を保持している。ただし、キャッシュメモリ部におけるディスクからの読み出しデータについては、マスタとスレーブ間で同一の内容を保持していなくても構わない。

障害が発生した場合の処理は、基本的に、実施例1と同様であるが、その概要を図18で簡単に説明する。マスタ／スレーブ キャッシュメモリ部、および、マスタ／スレーブリソース管理部は、障害監視機構により、相互に障害発生の有無について定期的な確認を行っている。障害監視機構によって発見された障害は、リソース管理部のバス制御機構に通知される。バス制御機構では、通知された障害情報の解析を行い、障害部位の特定を行う。バス制御機構では、障害部位が特定できれば、スイッチ内にあるフォワーディングテーブルを制御することで、障害部位を回避するようにバスの設定を行うことで、障害部位の切り離しが完了する。

本実施例では、専用のハートビート信号3を設けることにより、相互にその動作を確認するための障害監視機構Cを、実施例1より単純な構成で実現することができる。すなわち、ハートビート信号3を通して、お互いの動作状況を、直接、監視することができる。そのため、マスタ、もしくは、スレーブのいずれかのキャッシュメモリ部、もしくはリソース管理部に異常が発生した場合、その障害情報は、より迅速にもう一方のキャッシュメモリ部、もしくは、リソース管理部で検知することができる。

また、本実施例では、キャッシュメモリ部30内で検出された障害情報が、スイッチ40を介して(マスタ)リソース管理部50のバス制御機構Pに通知され、リソース管理部内のバス制御機構Pによってキャッシュメモリ部30の障害回復処理が行われる。これにより、障害情報をリソース管理部50に集めて、より適切な障害回復処理を行うことが可能となる。

また、本実施例では、ホストインターフェース側とディスクインターフェース側でスイッチを分離することにより、ホスト側とディスク側で柔軟な接続数変更が可能になるとともに、より大規模な構成への対応が可能となる。

【0025】

本実施例によれば、前述の実施例と同様、迅速、かつ、高信頼な障害回復処理を実現することができ、ストレージシステムの性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことがなく、また、障害のない通常の動作においての性能劣化は皆無で、かつ、低成本で高信頼に製造することが

可能である。

【0026】

《実施例3》

図4に、本発明の他の実施例を示す。

図4に示した実施例は、キャッシュメモリ部30、及び、リソース管理部50が、ハートビート信号3を備えていないことと、内部インターフェース信号のスイッチ網が冗長構成になっていることを除いて、実施例2の図3に示す構成と同様である。ただし、キャッシュメモリ部30は、障害監視機構Cとバス制御機構を実装していない。リソース管理部50は、障害監視機構C、バス制御機構Pとともに備えている。

本実施例では、キャッシュメモリ部30の障害監視も、リソース管理部50内の障害監視機構Cを用いて行う。その実現方法としては、リソース管理部の障害監視機構Cが、定期的にキャッシュメモリ部30へアクセスを行い、キャッシュメモリ部の動作状況を監視する方法や、ホストインターフェース部10およびディスクインターフェース部20からキャッシュメモリ部30へアクセスした際に障害が検出された場合、その障害情報をリソース管理部に報告する方法などが考えられる。また、本実施例では、各ホストインターフェース部、各ディスクインターフェース部に複数のポートを設け、スイッチも2重化することにより、ホストインターフェース部、および、ディスクインターフェース部とキャッシュメモリ部、もしくは、リソース管理部との間に複数のバスを設けてある。

これにより、リソース管理部、キャッシュメモリ部の機能障害だけでなく、リソース管理部、もしくは、キャッシュメモリ部とホストインターフェース部およびディスクインターフェース部との間のバス障害についても、障害回復を行えることになり、可用性をさらに高めることができる。

また、リソース管理部50に、障害監視機構とバス制御機構を集めることにより、より障害状況の的確な分析が行えることになり、適切かつ高信頼の障害回復処理が可能となる。

【0027】

本実施例によれば、前述の実施例と同様、迅速、かつ、高信頼な障害回復処理

を実現することができ、ストレージシステムの性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことがなく、また、障害のない通常の動作においての性能劣化は皆無で、かつ、低成本で高信頼に製造することが可能である。

【0028】

《実施例4》

図5に、本発明の他の実施例を示す。

図5に示した実施例は、キャッシュメモリ部30、及び、リソース管理部50が、ハートビート信号3を備えていないことと、複数のディスク制御サブユニット200を備えることを除いて、実施例2の図3に示す構成と同様である。複数のディスク制御サブユニット内のキャッシュメモリ部それぞれが、障害監視機構Cを備えている。リソース管理部50は、障害監視機構C、バス制御機構Pとともに備えている。

本実施例では、ディスク制御サブユニット200ごとに、キャッシュを分散配置することで、キャッシュの使用効率（ヒット率）を高めて性能を向上するとともに、ホスト側とディスク側で柔軟にシステム規模を拡張することが可能となり、より高スケーラブルなシステムの提供が可能となる。

また、本実施例では、実施例2と同様、キャッシュメモリ部30の障害に伴う障害回復処理も、リソース管理部50内のバス制御機構Pを用いて行う。実施例2、3と同様、リソース管理部50に、障害情報を集めることにより、より障害状況の的確な分析が行えることになり、ディスク制御サブユニット200の個数を増加させた、より大規模なディスク制御装置においても、適切かつ高信頼な障害回復処理が可能となる。

【0029】

本実施例によれば、前述の実施例と同様、迅速、かつ、高信頼な障害回復処理を実現することができ、ストレージシステムの性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことがなく、また、障害のない通常の動作においての性能劣化は皆無で、かつ、低成本で高信頼に製造することが可能である。

【0030】

《実施例5》

図15に、本発明の他の実施例を示す。

図15に示した実施例では、実施例1-4で記載のディスク制御装置が、複数のホストコンピュータとホストコンピュータ間ネットワークを介して接続し、複数の磁気ディスク装置と磁気ディスク装置間ネットワークを介して接続している。ホストコンピュータ間ネットワークには、ファイルシステムの処理を行うサーバ110（NASヘッド）、複数のディスク制御装置が管轄するストレージをまとめて管理するためのサーバ120（ディスク制御装置間仮想化エンジン）、データベース・インターフェース処理を行うためのサーバ130（DB機能付加エンジン）などが接続されることもある。これら、NASヘッド、仮想化エンジン、DB機能付加エンジンは、ディスク制御装置の中に実装しても構わない。

【0031】

本実施例によれば、迅速、かつ、高信頼な障害回復処理を行えるディスク制御装置を用いることにより、極めて可用性が高く、性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことのない、ストレージシステムの提供が可能となる。

【0032】

【発明の効果】

以上説明したように、本発明では、キッシュメモリ部30もしくはリソース管理部50の障害に対して、スイッチ40のフォワーディング・テーブルとキッシュメモリ部30もしくはリソース管理部50のLIDの変更のみで、障害部位の切り離しが完了し、従来技術のように、複数のホストインターフェース部10及びディスクインターフェース部20の間でのブロードキャスト通信や複雑な制御は不要である。そのため、障害発生時においても、迅速、かつ、高信頼な障害回復処理を実現することができ、ストレージシステムの性能劣化や、ホストコンピュータ上のアプリケーション動作不良を引き起こすことがない。

また、本発明において、スイッチ内フォワーディング・テーブルは、障害発生時にのみ変更されものであって、従来技術のように、通信を行うたびにパケット

の宛先の解釈や変更が行われる複雑なスイッチを必要としない。そのため、障害のない通常の動作においての性能劣化は皆無で、かつ、低コストで高信頼に製造することが可能である。

【0033】

本発明では、障害監視機構による障害発生通知を、パス制御機構で解析し、フォワーディングテーブルを制御するので、柔軟なシステム構成に対応可能である。とくに、複数のディスク制御サブユニットが存在する大規模なディスク制御装置の場合でも、複数の障害監視機構からの障害情報をパス制御機構に集めることにより、より障害状況の的確な分析が行え、高信頼な障害回復処理が可能となる

【図面の簡単な説明】

【図1】

本発明によるディスク制御装置における障害回復処理の動作原理を示す図である。

【図2】

本発明によるディスク制御装置の構成を示す図である。

【図3】

本発明によるディスク制御装置の構成を示す図である。

【図4】

本発明によるディスク制御装置の構成を示す図である。

【図5】

本発明によるディスク制御装置の構成を示す図である。

【図6】

本発明によるディスク制御装置内のスイッチの構成を示す図である。

【図7】

本発明によるディスク制御装置内のキャッシュメモリ部の構成を示す図である

【図8】

本発明によるディスク制御装置内のホストインターフェース部の構成を示す図である。

【図9】

従来のディスク制御装置の構成を示す図である。

【図10】

従来のディスク制御装置内のスイッチの構成を示す図である。

【図11】

従来のディスク制御装置内のキャッシュメモリ部の構成を示す図である。

【図12】

従来のディスク制御装置内のホストインターフェース部の構成を示す図である

【図13】

本発明の障害監視機構の動作を示す図である。

【図14】

本発明のバス制御機構の動作を示す図である。

【図15】

本発明のディスク制御装置を用いたストレージシステムの例を示す図である。

【図16】

本発明のディスク制御装置においてホストコンピュータからキャッシュメモリ部へのコマンド送信の概要を示す図である。

【図17】

本発明のディスク制御装置の構成を示す図である。

【図18】

本発明のディスク制御装置の障害処理の概要を示す図である。

【符号の説明】

- 1 ホストインターフェース信号
- 2 ディスクインターフェース信号
- 3、5 ハートビート信号
- 4 内部インターフェース信号
- 6 データインターフェース信号
- 7 管理インターフェース信号

10、1X ホストインターフェース部

11 DMAエンジン

12 送信データインターフェース

13 受信データインターフェース

14 プロトコル制御プロセッサ

15 ローカルメモリ

16 バスインターフェース

17 プロセッサインターフェース

20、2X ディスクインターフェース部

30、30A、30B、3X キャッシュメモリ部

31X データバスインターフェース

32X ヘッダ解析部

33 パケットバッファ

34 DMAエンジン

35 メモリ制御回路

36 プロセッサ

37 キャッシュメモリ

38 セレクタ

39 調停回路

40、40A、40B、4X スイッチ

41 パスインターフェース (PORT)

41X データバスインターフェース

42、42X ヘッダ解析部

43 パケットバッファ

44 アドレスラッチ

45 経路変更手段

46、46A、46B フォワーディング・テーブル

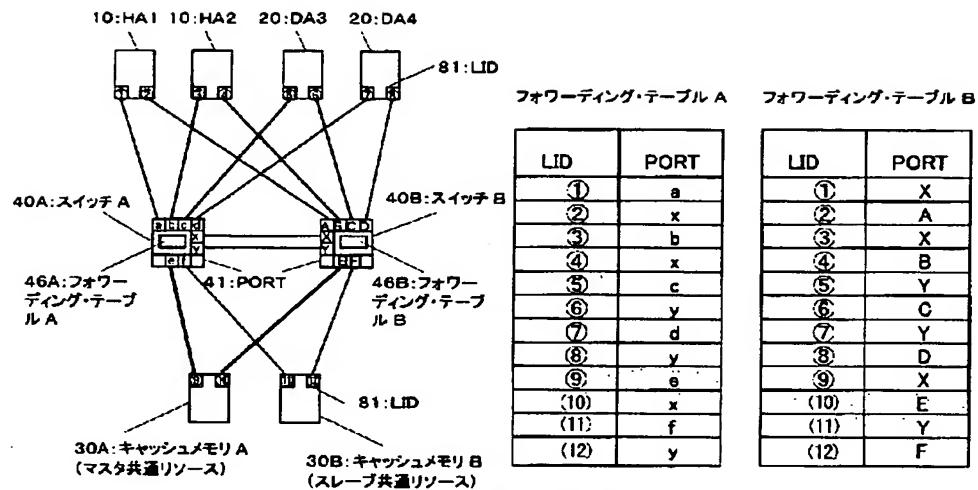
47 セレクタ制御信号

48 セレクタ

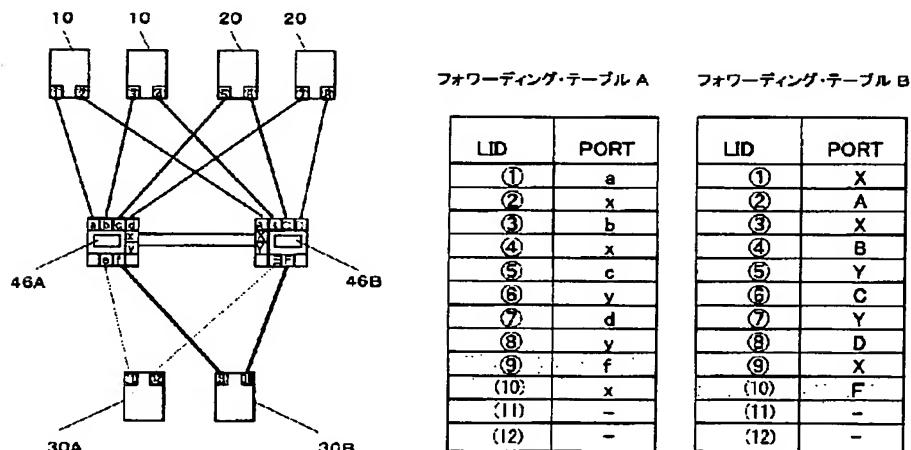
50 リソース管理部
60 ホストコンピュータ
70 磁気ディスク装置
80、8X 内部プロトコル処理部
81 ローカルID情報(LID)
82 送信PHY
83 受信PHY
84 プロトコル制御プロセッサ
85 ローカルメモリ
86 バッファ
87 リンク処理
88 ヘッダ制御
90 チャネルプロトコル処理部
92 送信PHY
93 受信PHY
96 バッファ
97 リンク処理
98 トランスポート処理
100～104 ディスク制御装置
110 NASヘッド
120 ディスク制御装置間仮想化エンジン
130 DB機能付加エンジン
200 ディスク制御サブユニット
C 障害監視機構
P パス制御機構。

【書類名】 図面
【図1】

図1



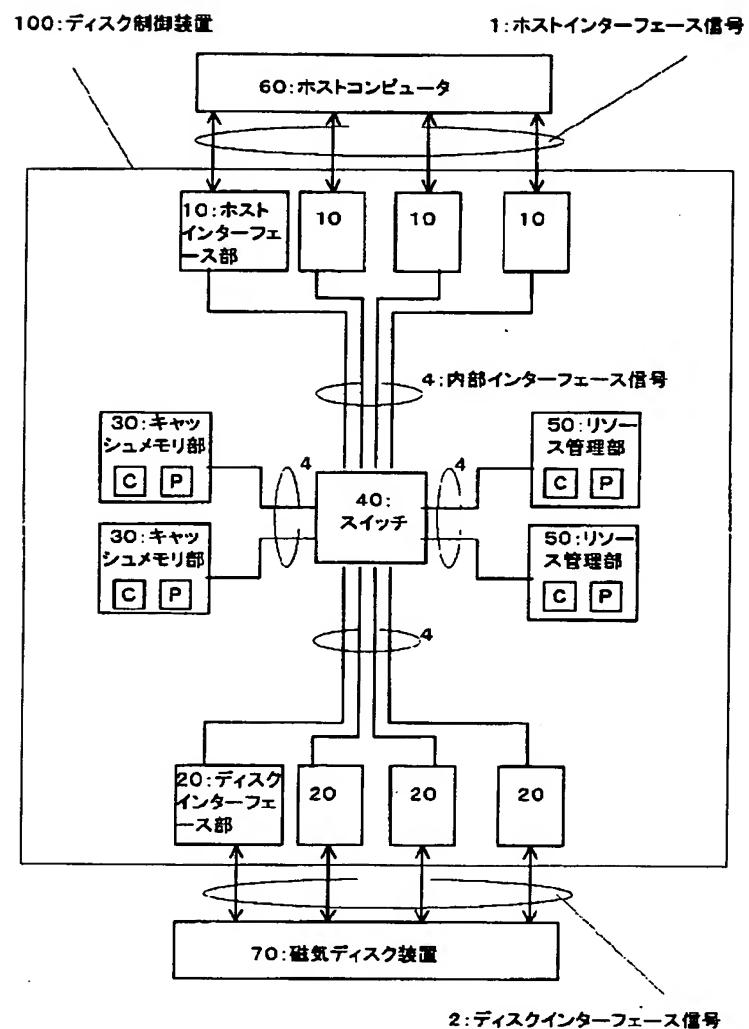
(1)障害前の状態



(2)30Aの障害処理後の状態

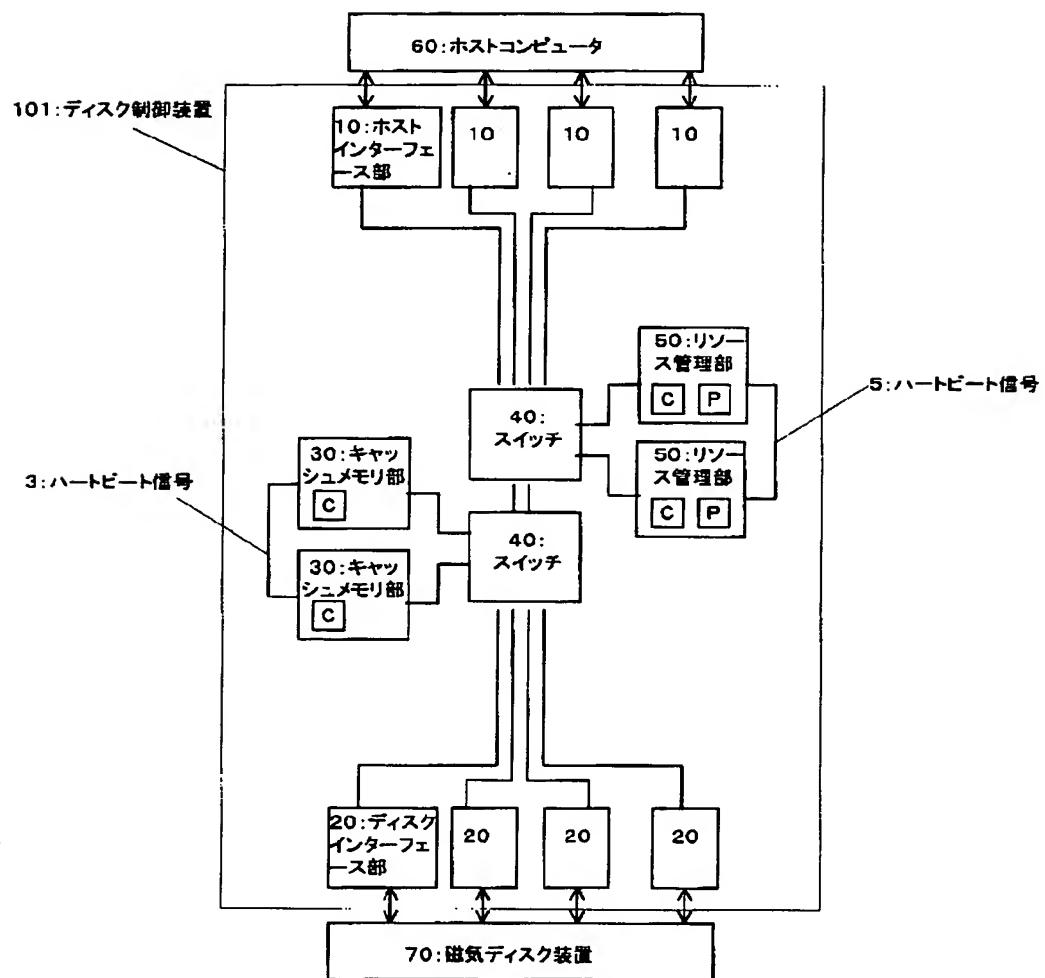
【図2】

図2



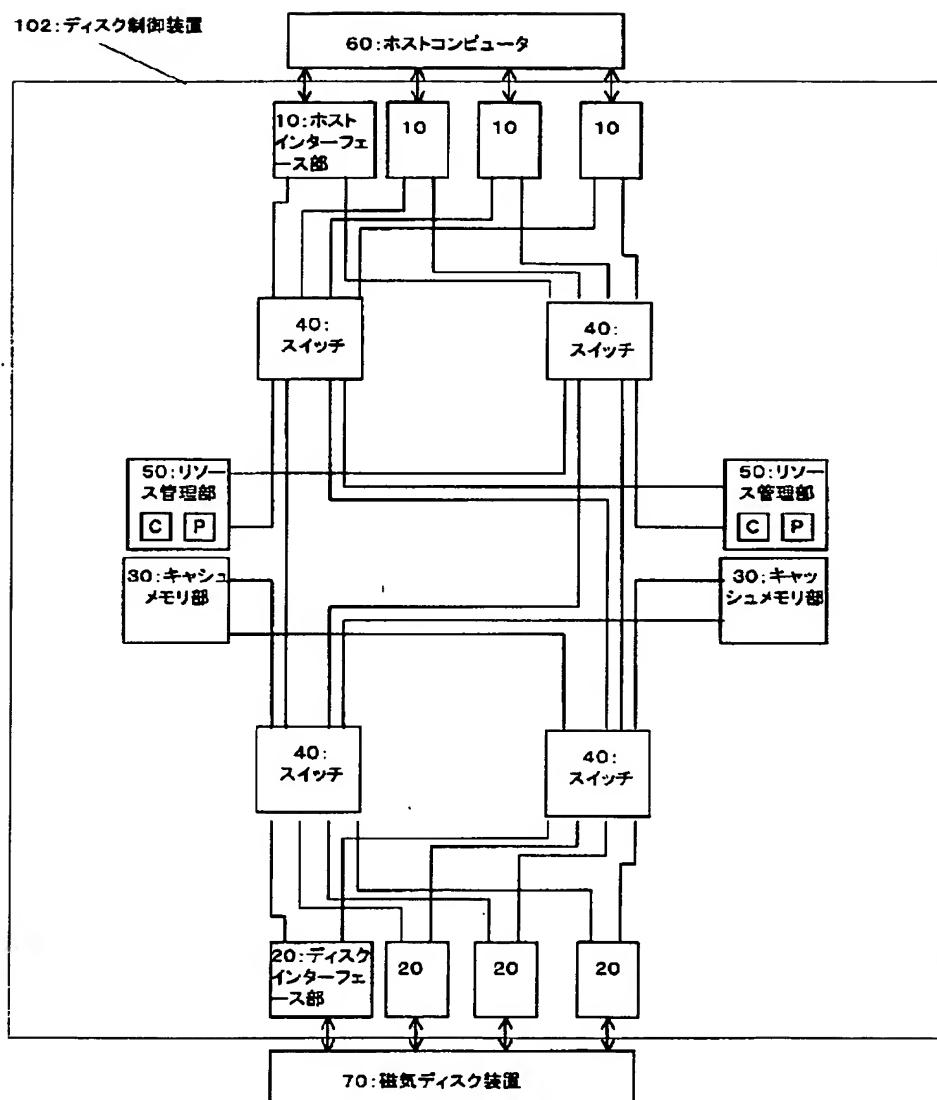
【図3】

図3



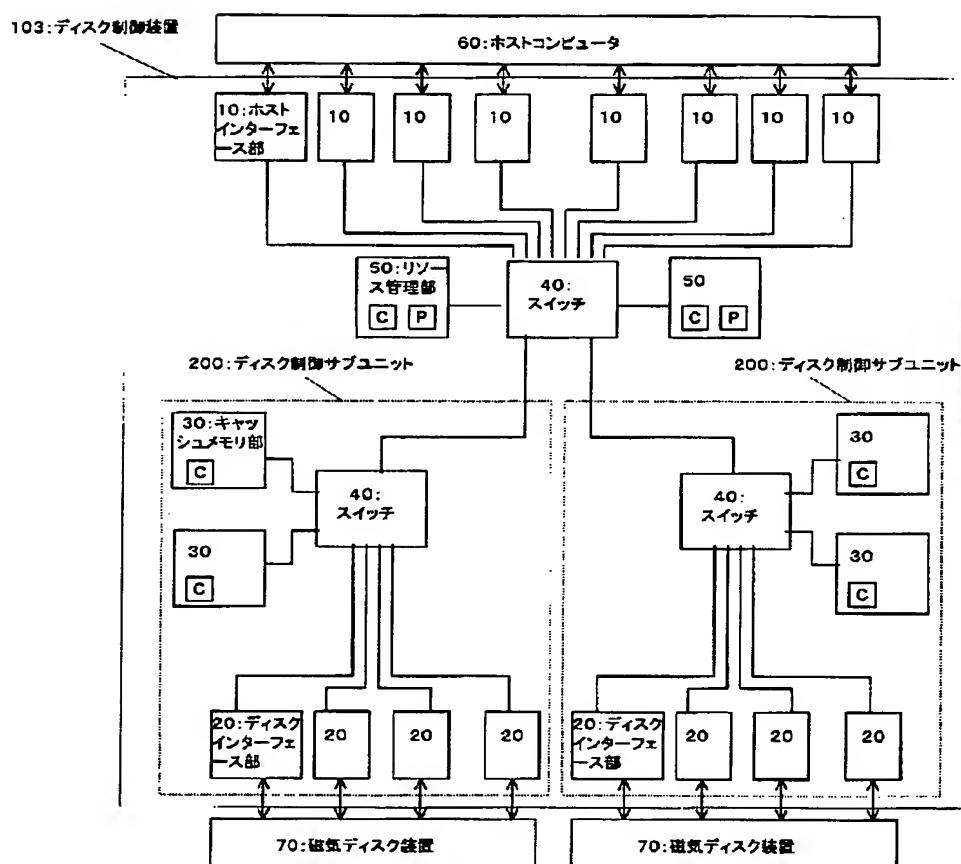
【図4】

図4



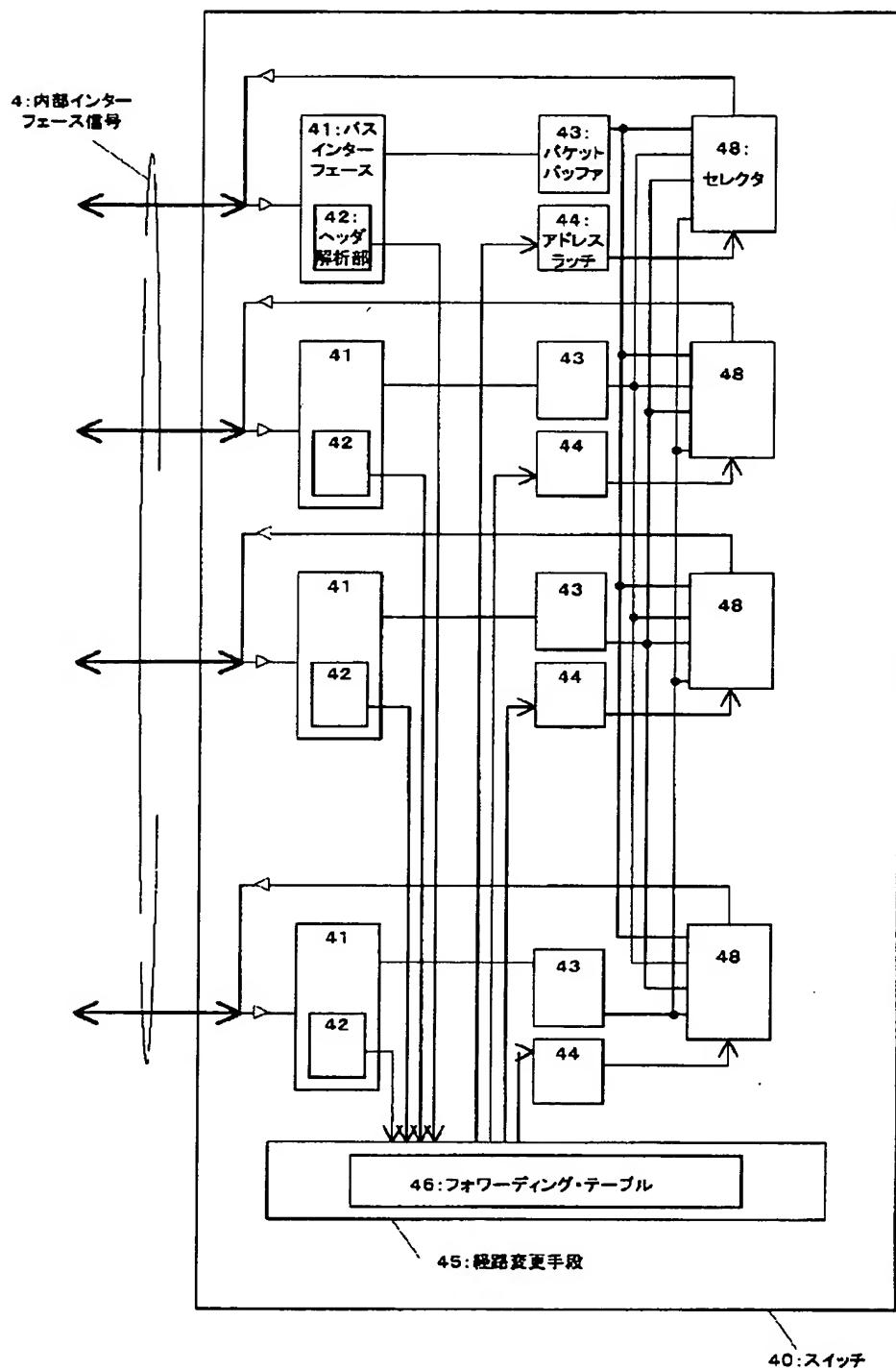
【図5】

図5



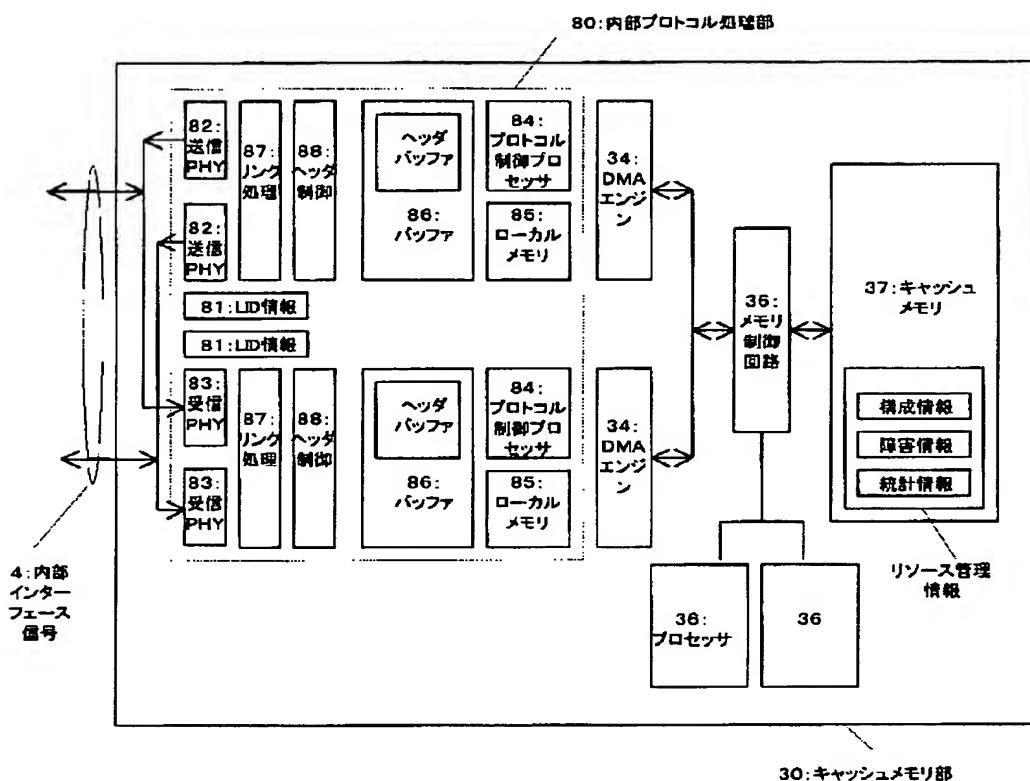
【図6】

図6



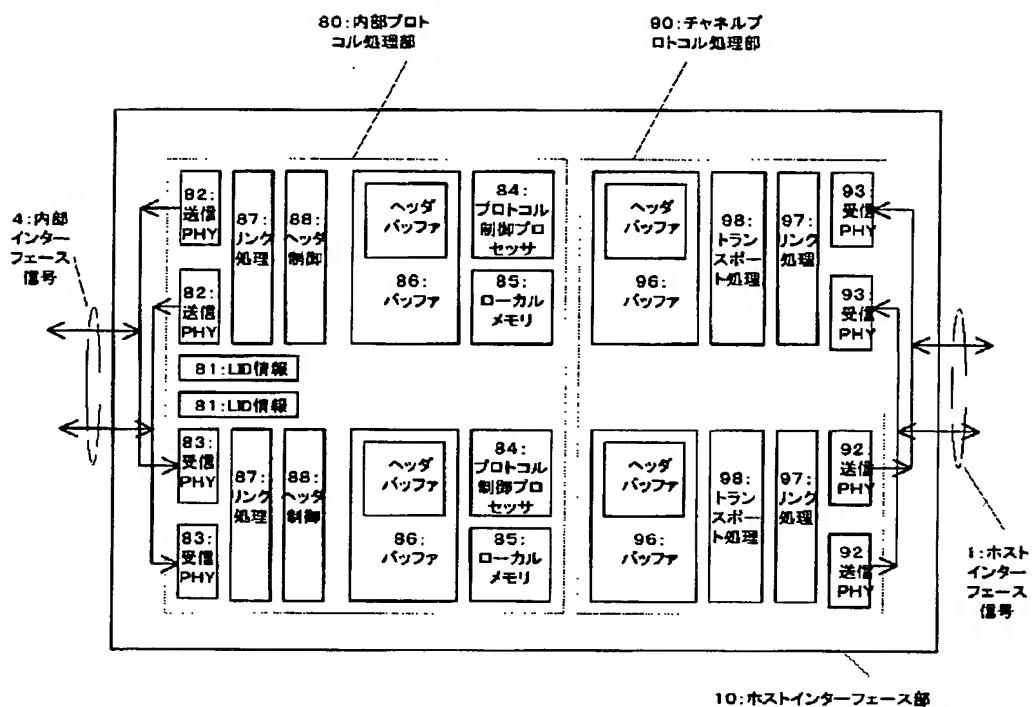
【図7】

図7



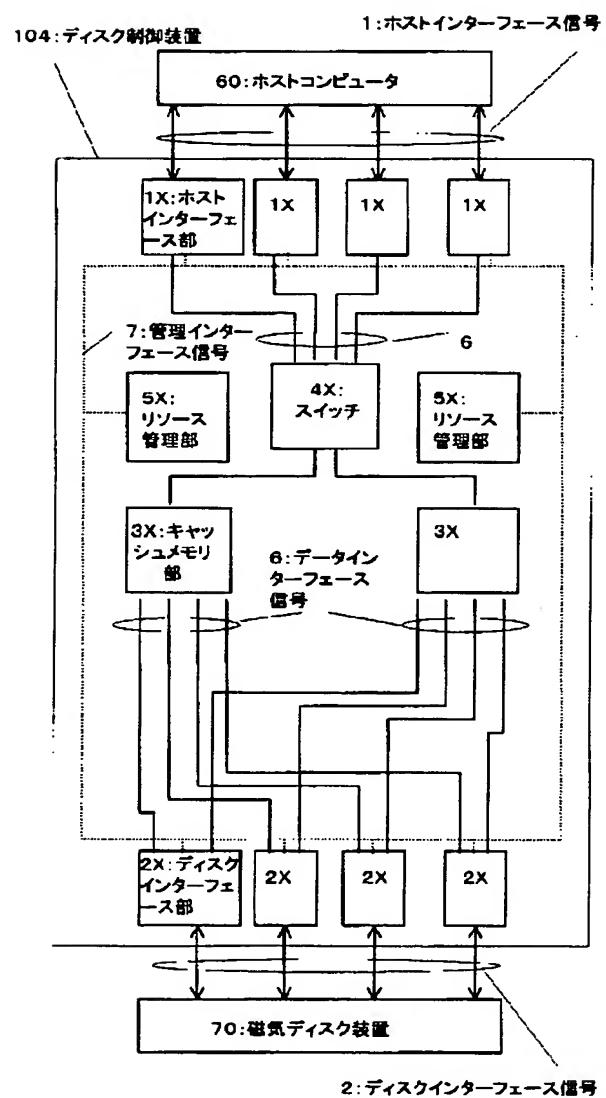
【図8】

図8



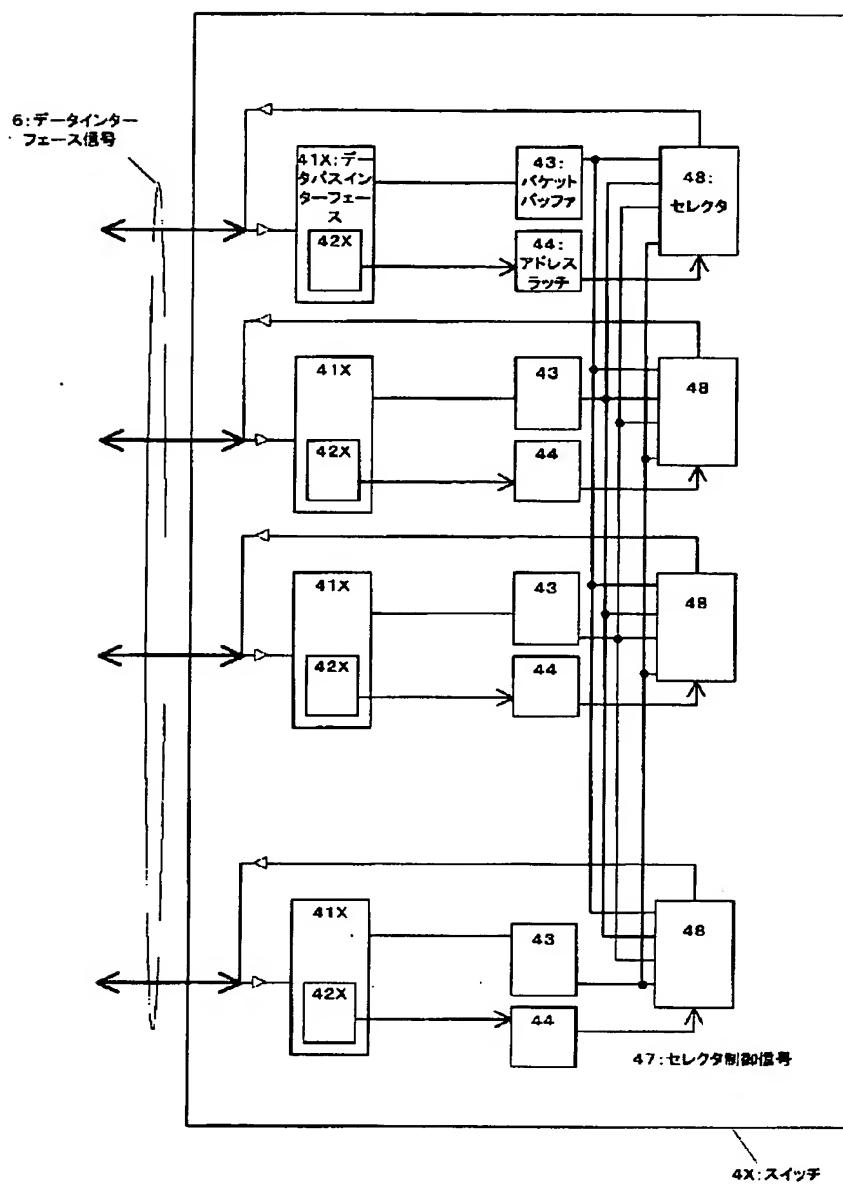
【図9】

図9



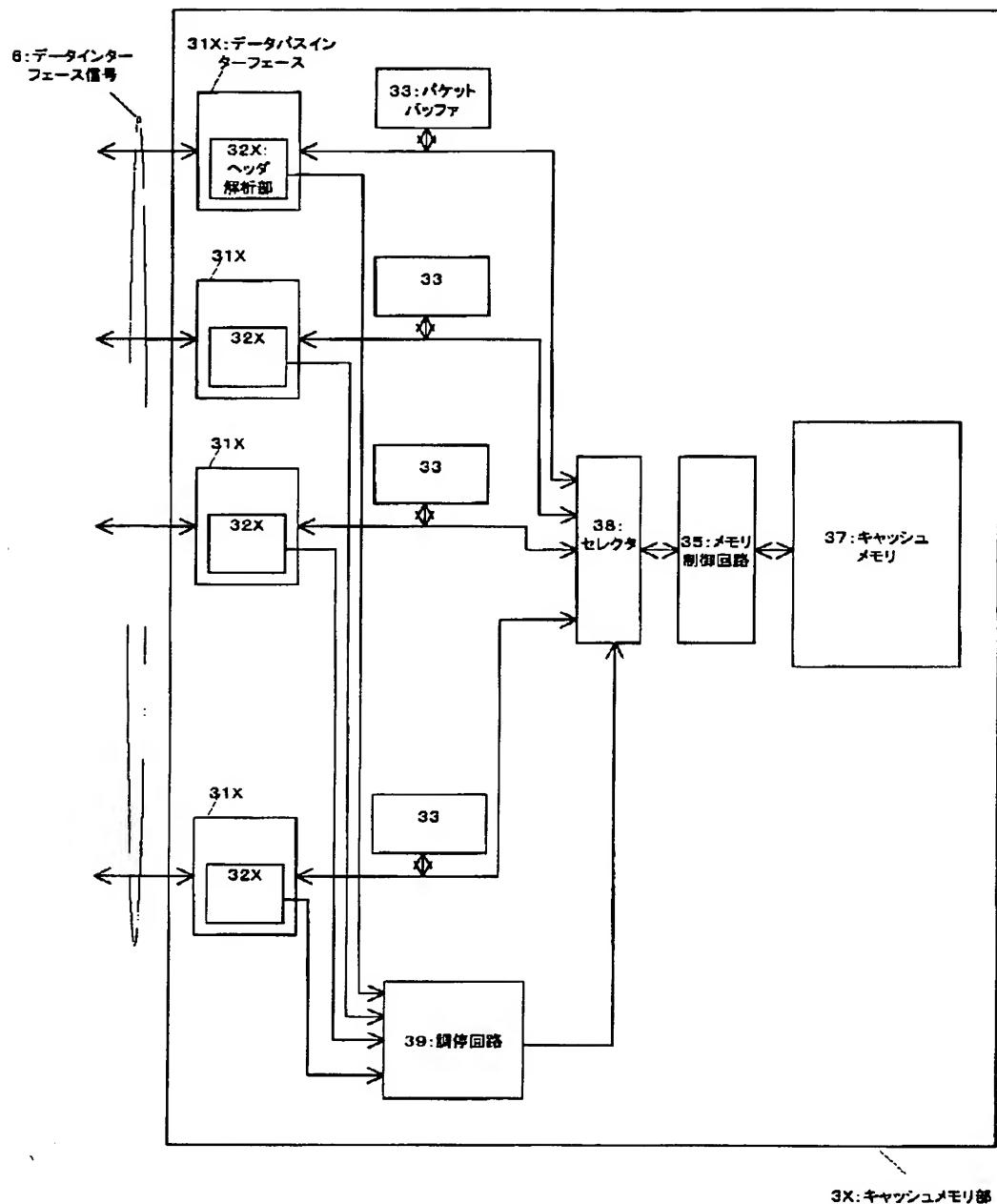
【図10】

図10



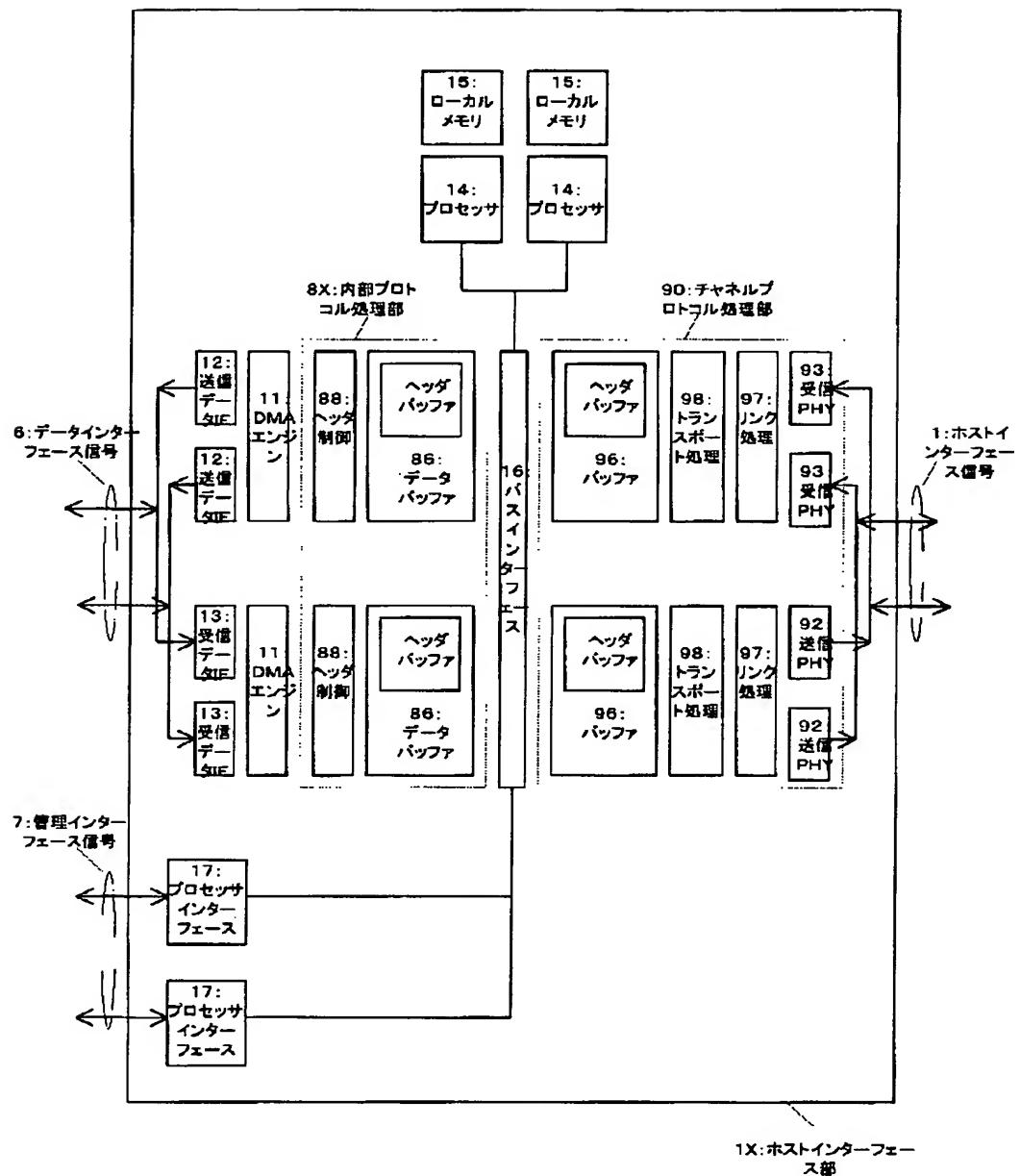
【図11】

図11



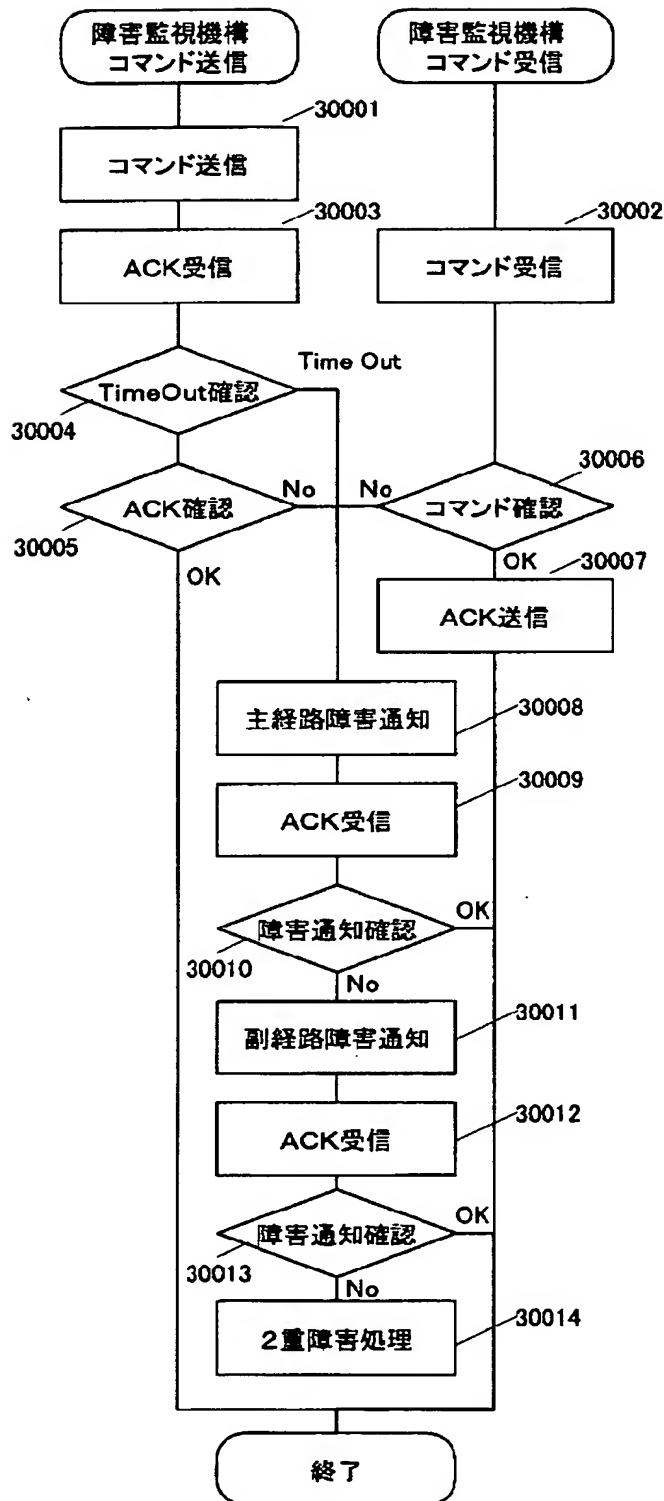
【図12】

図12



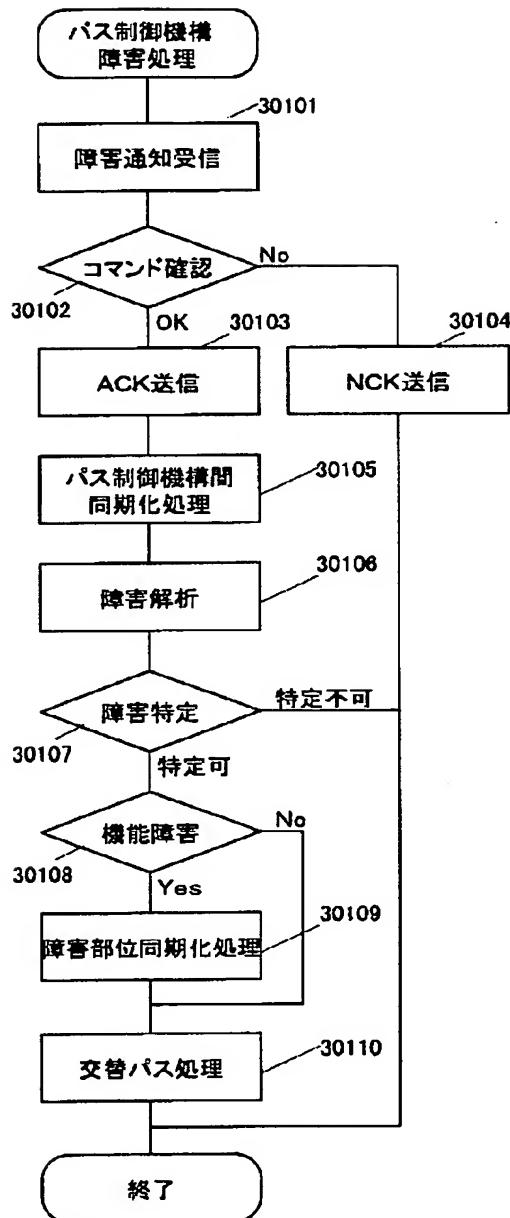
【図13】

図13



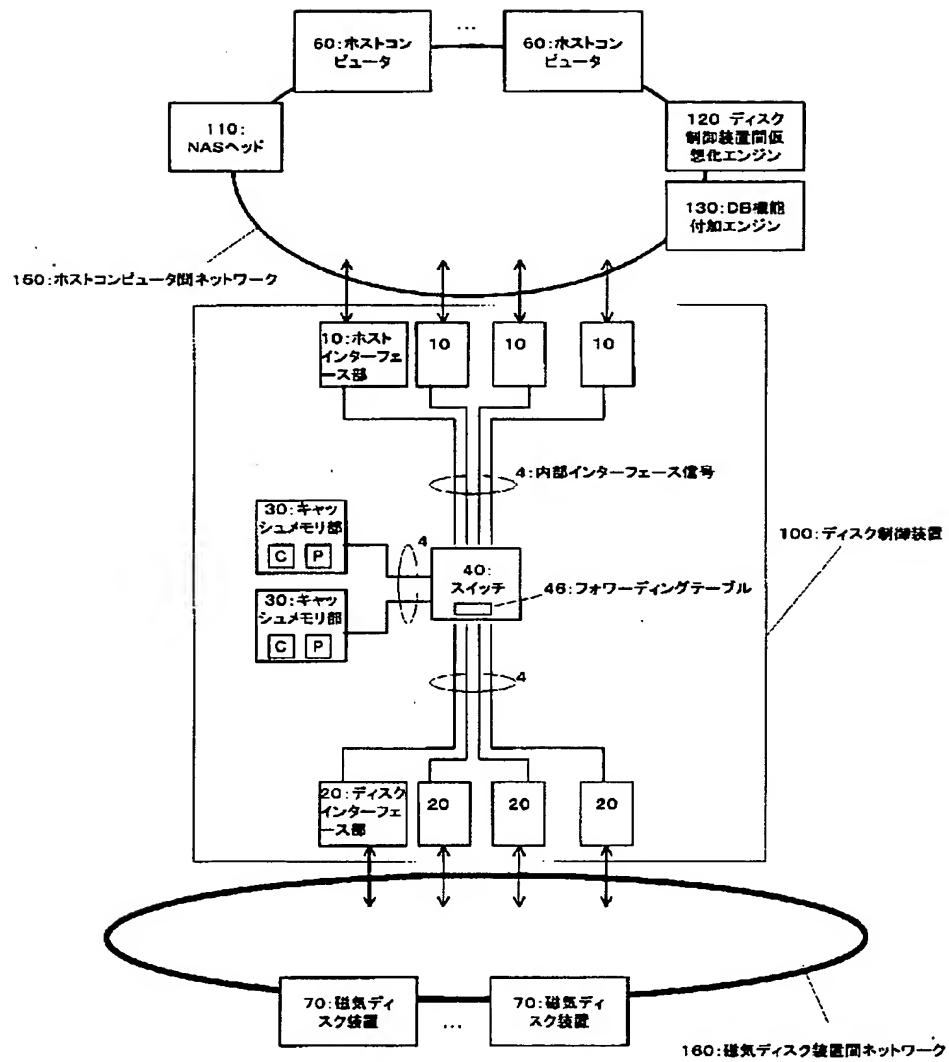
【図14】

図14



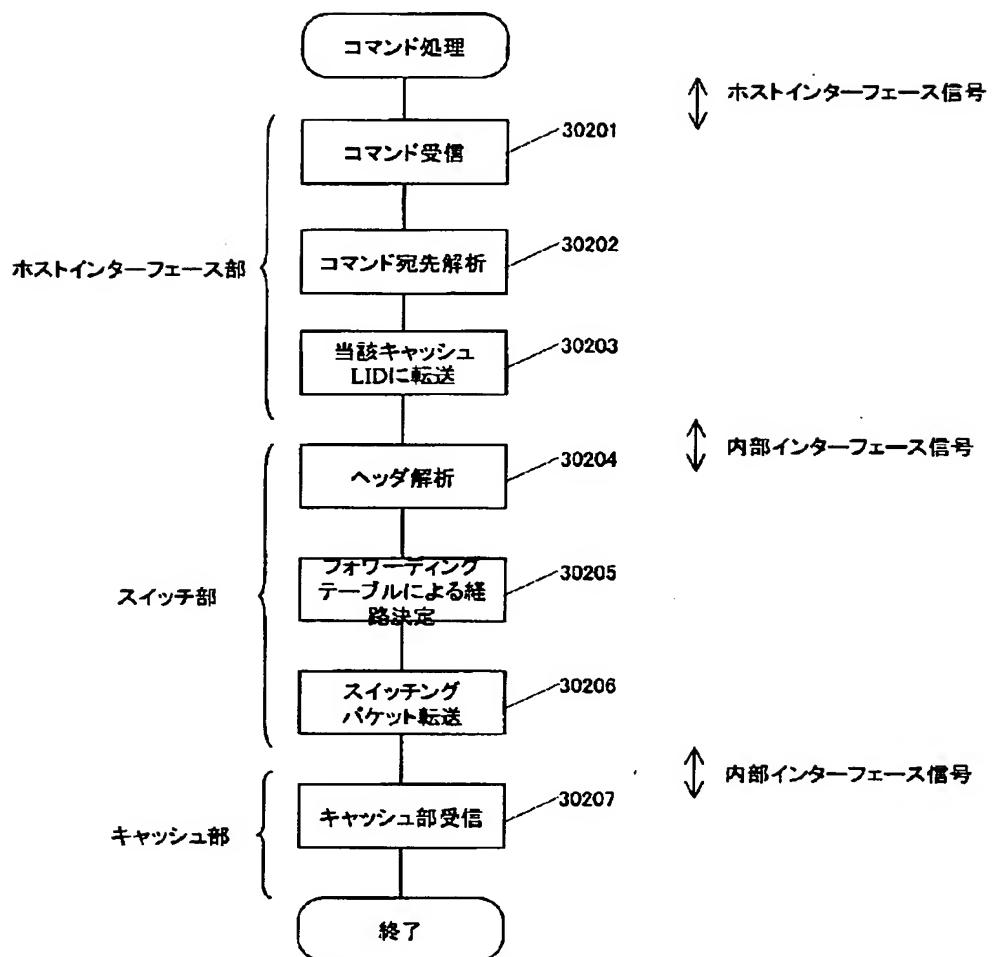
【図15】

図15



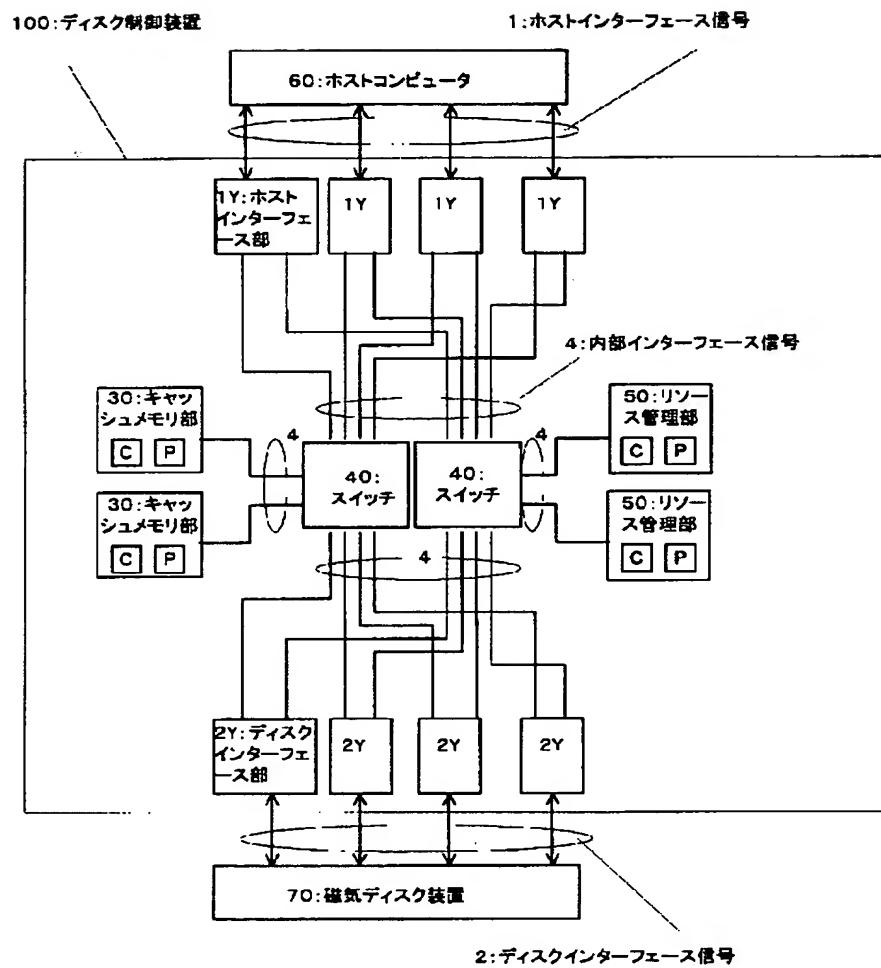
【図16】

図16



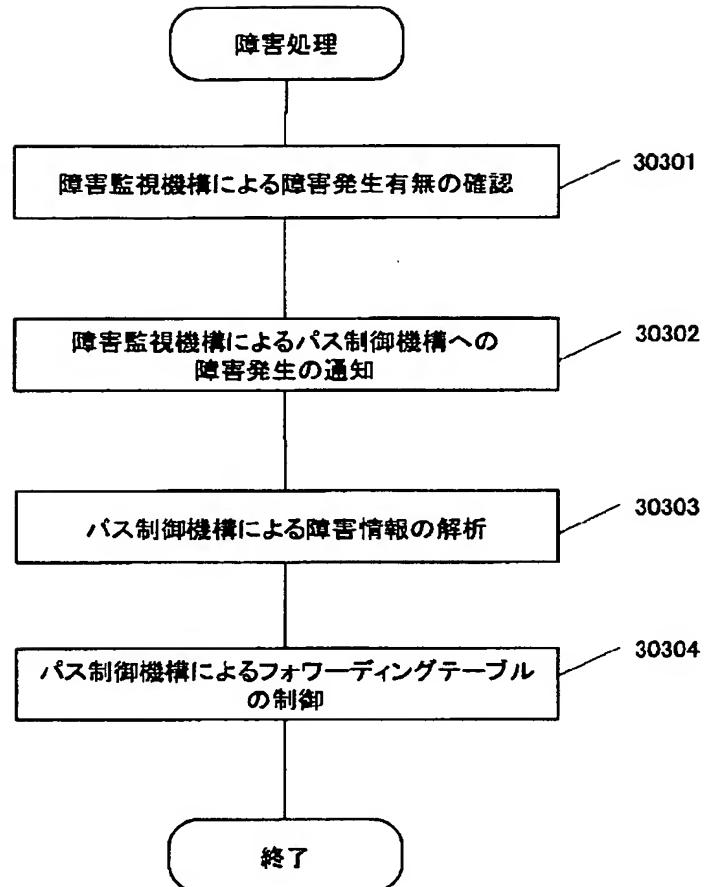
【図17】

図17



【図18】

図18



【書類名】 要約書

【要約】

【課題】 ストレージ・システムの性能劣化や、ホスト・アプリケーションの動作不良を引き起こすことのない高可用性ディスク制御装置を提供する。

【解決手段】 複数のホストインターフェース部及び複数のディスクインターフェース部とキャッシュメモリ部との間を、1つ以上のスイッチで構成されるスイッチ網を介して接続し、スイッチにスイッチ網内の経路を指定するためのフォワーディングテーブルと該テーブルを変更する変更手段を設け、また、ホストインターフェース部、ディスクインターフェース部及びキャッシュメモリ部に、スイッチ網内で一意に決まるローカルIDと該IDを変更する変更手段を設け、さらに複数のキャッシュメモリ部に、その障害発生の有無を監視するための障害監視機構と障害発生時に障害部位を回避するように前記スイッチ内のフォワーディングテーブルを制御するためのパス制御機構を設ける。

【選択図】 図1

認定・付加情報

特許出願の番号	特願2002-378956
受付番号	50201981801
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 1月 6日

<認定情報・付加情報>

【提出日】 平成14年12月27日

次頁無

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所